Interface, multimodal / multisensory

Joan de Boeck, Giovanna Varni, Armen Khatchatourov

Interface, multimodal / multisensory

Joan De Boeck [EDM, UHASSELT]
Giovanna Varni [InfoMus Lab]
Armen Khatchatourov [COSTECH]

When one speaks about a multimodal interface, it refers to the second definition of multimodality [→ Multimodality, in human-computer interaction]: a multimodal interface is an interface that supports various means for the user to express or interpret information e.g.: keyboard, mouse, spoken language, icons, etc. In general, a multimodal interface is a class of interfaces, designed to make the interaction process between a human and a computer more similar to human-to-human communication (e.g., by means speech, gesture, emotions), by exploiting the human's natural use multimodal interaction. Users can interact with this type of interfaces in a more natural and transparent way by the integration of multiple input and output modes.

What is important in a multimodal interface, is that these kinds of systems strive for meaning [Nigay and Coutaz, 1993], as defined in multimodality from the point of view of HCI. If we consider a system-centered view, a multimodal system has the capacity to communicate with a user along different types of communication channels –and- to extract and convey meaning automatically. A system that allows the user to involve various action-perception, focusing solely on the sensory aspect, as defined in multimodality from the point of view of psychology, would rather be called a multisensory or multimedia system. We may observe that both multimedia and multimodal systems use multiple communication channels. But in addition, a multimodal system is able to automatically model the content of the information at a much higher level of abstraction.

The relationships between the different modalities, as well as the user's preference can be expressed by means of the CARE properties [Coutaz, 1995].

- Complementary:

Modalities are complementary when all the modalities are necessary for completing the task, but each is carrying just a part of the information. A typical example is a spoken command that must be accompanied by a pointing gesture, to indicate the subject of the command.

- Assignment:

A modality is assigned if there is no other modality to execute the task.

- Redundancy

Modalities are redundant if they have the same expressive power for the task (see equivalence) but all of them must be used.

- Equivalence:

Modalities are equivalent for completing a task if it is necessary and sufficient to choose one of them.

All relations can be permanent or transient and are total or partial. A relation is permanent if it is true in any state of the application, otherwise it is transient. A relation is total if it applies to every task of the application, otherwise it is partial.

It may be clear that in a decent multimodal interface, the system CARE [Coutaz, 1995] properties must meet the user's preference (user CARE properties), while for consistency, the relationship must be permanent and total.

In the scope of the research on enactive interfaces, an important question is how are they positioned with regard to multimodal interfaces, and how both are related to each other.

As a multimodal interface focuses on exchanging meaning by means of multiple sensory channels, and enactive interfaces focus on our human knowledge acquired by (multisensory feedback while) doing, one can state that enactive interfaces are particular and more specific kind of multimodal interface, that besides the striving for meaning,

also allows users to learn from their sensory-motor actions.

Another way to distinguish between multimodal and enactive interfaces is the following consideration: As referred in [Oviatt, 2002], user(s) can interact with this type of interfaces in a natural and transparent way by means the integration of multiple active and/or passive input modes. "*Active input modes are ones that are deployed by the user intentionally as an explicit command to a computer system*", whereas "*passive input modes refer to naturally occurring user behaviour or actions that are recognized by a computer. They involve user input that is unobtrusively and passively monitored, without requiring any explicit command to a computer*". Example of the active input modes is speech, examples of passive input modes can be considered facial expression, gaze and manual gestures. The integration of modalities is the central issue to be faced by these interfaces. In multimodal interfaces, both modes (explicit command and recognized passive command) are used to control the system's behaviour in an explicit way (the recognized command is interpreted by the system in a univocal way). In multimodal interfaces, the emphasis is put not on the control, but on the interaction, and there may not be the recognition at all [→ Mapping and control vs. instrumental interaction]

References

[Coutaz, 1995] Joelle Coutaz, Laurence Nigay et al, Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties in Proceedings of INTERACT95, 1995

[Oviatt, 2002] S. Oviatt, in Handbook of Human-Computer Interaction, (ed. by J. Jacko & A. Sears), Lawrence Erlbaum: New Jersey, 2002.

Related items

Classification of perceptual modalities
Interface
Interface, enactive
Interface, ergotic
Mapping and control vs. instrumental interaction
Multimodality, in human-computer interaction